DOCUMENT RESUME

ED 038 656 CG 005 107

AUTHOR	Blalock, H. M.
TITLE	The Future of Sociological Research: Measurement Errors and Their Implications.
INSTITUTION	North Carolina Univ., Chapel Hill.
PUB DATE	[70]
NOTE	38p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.00
DESCRIPTORS	Data Analysis, Data Collection, *Evaluation Criteria, *Measurement Techniques, *Research Methodology, *Research Problems, *Sociology

ABSTRACT

The report deals with the relationship between measurement and data analysis procedures in sociological research. The author finds that too many measured variables exist in both theory and measurement assumptions. Since these procedures are interrelated, improvements in either or both areas are necessary. Presented are three sections: (1) specific complications and distortions which may be produced by random and nonrandom measurement errors; (2) statistical approaches designed to eliminate or estimate the effects of measurement errors; and (3) exploration of some practical implications for the profession. The first section includes discussions on multicollinearity, choosing among alternative causal models, and nonlinearity and nonadditivity. The second section deals with various approaches of estimating random and nonrandom errors. The third section concludes that although much remains to be accomplished research sociologists should not be immobilized because of well-recognized but unmet assumptions. (Author/MC)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED038656

THE FUTURE OF SOCIOLOGICAL RESEARCH: MEASUREMENT ERRORS AND THEIR IMPLICATIONS

H. M. BLALOCK, JR.

UNIVERSITY OF NORTH CAROLINA

Let me begin with an obvious point. The future of data analysis in sociological research depends on the quality of our theories and of our measurement procedures. The fact that data analysis is today something of an art that requires numerous distinct judgmental decisions, each with potentially different implications, is a reflection of the fact that there are numerous competing theoretical explanations for each piece of information at our disposal. Likewise, there is no real consensus on our measurement procedures or definitive criteria that can be used to decide among indicators of a given concept. The extent and nature of measurement errors are often unknowns, and we have yet to develop the practice of attempting to estimate measurement errors and to incorporate these estimated errors into our tests. In view of all this, it is indeed surprising to find two or more analysts reaching essentially the same conclusions when confronted with the same data.

To the degree that a theory is clearly specified and its implications made fully explicit, and to the degree that our measurement problems are resolved, we can expect to find data analysis becoming increasingly routine and implied by the theory and measurement procedures. Conversely, to the degree that the theory and measurements are not specified, a heavier burden will fall on the data analyst to generate reasonable theoretical explanations as a product of his explorations with the data. But in doing so, he

CG005107

will also have to make assumptions about the adequacy of his measurement procedures and the linkages between his indicators and the underlying variables of real interest.

The task of the methodologist is to explicate this process of interrelating theory, measurement, and data analysis procedures. I believe we have made very good progress during the decade of the 1960's in learning more about the logic of data analysis and the sophisticated use of multivariate analysis procedures. Much of this is due to our having been able to borrow ideas from more advanced fields. However, it has presumed the existence of reasonably adequate theories and the nonexistence of measurement errors. Since it is far less likely that we will be able to borrow our theories and measures from these other disciplines, we will henceforth be much more on our own. I think we have reached a point where our knowledge of how to analyze data has outstripped our theoretical knowledge and our present measurement capabilities and that we must therefore turn our energies during the decades of the 1970's and 1980's more in the direction of theory construction and measurement. Unless we do so, I cannot predict a very bright future for data analysis. For it is obvious that any form of analysis aimed at genuine explanations of social phenomena requires a priori theoretical assumptions about measurement errors.

These points are all very general and probably not very controversial. I should like to make them much more concrete by discussing the relationship between measurement and data analysis, putting aside the equally complex problems of the relationship between theory and measurement, and between theory and analysis. Testing a theory obviously requires a supplementary or auxiliary theory concerning the linkages between the variables of theoretical interest and one's indicators, as well as the disturbing influences

that may also affect the data.[1] This auxiliary theory may be of no real interest substantively, and may be of a rather mundane sort. But it nevertheless cannot be neglected if we are to adhere to any reasonable standards as to what we mean by "testing" a theory. It seems to me that we have thus far been satisfied with extremely crude tests of our theories. If a theory predicts the direction of a few total correlations, and if results are significant at the .05 significance level, then the theory has been "tested." But what about the numerous alternative theories, including those that allow for measurement errors? The fact that tests of the null hypothesis enable us to rule out the simple "chance" alternative does not mean very much, given the wide variety of spurious relationships and other possible kinds of distortions.

The causal approach to measurement errors, which involves the inclusion of both unmeasured and measured variables in explicit causal models, has given us a number of important insights into the ways in which random and nonrandom measurement errors can be conceptualized, the kinds of complications they produce, and the limitations of statistical approaches to estimating these errors. In the following two sections I shall discuss nontechnically a number of specific conclusions that have been reached. If nothing else, these results should sensitize sociologists to the importance of careful attention to measurement problems at the data collection stages of the research process.

In very general terms, the presence of unmeasured variables and measurement errors in a causal system introduces more unknowns than can be handled unless additional a priori assumptions are made. The more faith we have in the underlying theory the more such assumptions we can make in order to estimate our measurement errors. And the more faith we have in our measurement

assumptions, the simpler our auxiliary theory can be, and the better position we will be in to test complex theories and to obtain accurate estimates of the fundamental causal parameters. But when both the theory and the measurement assumptions are in doubt, we shall be in serious trouble.

Of course this is precisely the position we are in with respect to contemporary sociology: we do not really believe our theories nor do we have much faith in our measurements. We may blindly accept the adequacy of the one in order to assess the other (as we do when we assume measurement errors to be negligible), but this does not really help. In effect, our theories and measurement procedures are inextricably bound together in the testing and estimating phases of research, and improvements in the one can only come about through improvements in the other. However, since numerous subleties seem to be involved in the process, a good deal more methodological research will be necessary before we can adequately formulate the problem in such a way that we can deal with it in a cumulative fashion.

In the next section we shall examine a few specific kinds of complications and distortions that may be produced by random and nonrandom measurement errors. In the following section we shall then deal with statistical approaches designed to eliminate or estimate these effects of measurement errors. The final section will explore some practical implications for the profession.

DISTORTIONS PRODUCED BY MEASUREMENT ERRORS

Multicollinearity. In nonexperimental research we commonly find that many of the independent variables, the effects of which we wish to isolate, are themselves more highly intercorrelated than each is to the dependent

variables under study. As we develop relatively complex causal models of real processes, I believe we shall necessarily utilize what have been referred to as "block-recursive" systems, an example of which is given in Figure 1.[2] The essential feature of such systems is that variables are assumed to form blocks or sets in such a way that, although there may be reciprocal causation or feedback within blocks, there is one-way causation or negligible feedback between blocks. In fact, it can be shown that a block-recursive model must be assumed in order to delimit the variables to any finite number.[3] All variables among which there is reciprocal causation must be analyzed in terms of simultaneous equations that cannot legitimately be dealt with separately. However, it is possible to omit variables in all higher-numbered blocks that are taken to be dependent upon the block of variables under consideration, and variables in lower-numbered blocks may be selectively introduced for the purpose of solving for the unknowns in the system.

For example, suppose variables in block 1 are systemic variables which change only very slowly and which are taken as "situational factors" that influence or set limits to behavior. Perhaps block 2 variables are "background" factors relating to the early socialization of an individual, whereas variables in block 3 are his present status characteristics. Block 4 variables may constitute a set of present attitudes, whereas those in block 5 may be behavioral variables that the theory is designed to explain. If we were primarily interested in explaining block 3 status variables we could then neglect variables in blocks 4 and 5, but we would want to include variables in blocks 1 and 2. If we wanted to allow for major feedback from block 5 to block 3, then we would not be justified in analyzing blocks 3-5 individually.

Very frequently we wish to assess the relative effects of variables within the same block on certain specific variables in a higher-numbered block. For example, we may wish to compare the effects of status variables such as education, occupation, and income on attitudes, or those of various attitudinal variables on behavior. In such instances we are likely to find higher intercorrelations among the several independent variables than between these variables and the dependent variable. In another very common kind of situation, we wish to assess whether variables in one particular block have greater explanatory power than variables in a second block, which is also correlated with the first block.

In all of these instances it will be necessary to obtain accurate measures of each variable, and the need for accuracy will increase with the intercorrelations among the independent variables. Let us see why this is the case. It turns out that purely random measurement errors in any particular independent variable will attenuate both the correlation with the dependent variable and the regression coefficient linking these variables. If we take the measured value of X as X' and assume this is related to X by the simple formula $X' = X + e$, where e is a random variable (and therefore uncorrelated with X), then $\sigma^2_{X'} = \sigma^2_X + \sigma^2_e$, meaning that the variance of the measured value of X will be greater than the true variance. Furthermore, the amount of attenuation in the slope estimate (and also the correlation) increases as σ^2_X becomes small relative to the measurement error variance. An approximate formula for the slope attenuation in the bivariate case is given by

$$E(b_{YX'}) = \frac{\beta_{YX}}{1 + \sigma^2_e/\sigma^2_X} = \frac{\sigma^2_X}{\sigma^2_{X'}} \beta_{YX}$$

where β_{YX} represents the true slope and where $E(b_{YX'})$ represents the expected

value of the least squares estimate based on the measured value X'.[4]

If the equation for Y contains several intercorrelated independent variables, then random measurement error in one of them will ordinarily make the others look better. That is, although the slope estimate of the poorly measured variable will be biased downward, the slopes of the remaining variables will be biased upward. In effect, they will receive some of the credit that should have been awarded to the poorly measured variable. Of course more than one variable may be measured inaccurately, but the same general principle holds.

Gordon has constructed a number of different data sets to illustrate this principle.[5] In one such set, diagrammed in Figure 2, there are two blocks of independent variables, with all intercorrelations within blocks being .30, and with each possible correlation between variables in different blocks being .20. The first block contains three variables and the second only two. Two sets of results involving slightly different correlations with the dependent variable Y are given below:

		Set A		Set B	
		Correlation with Y	Partial Slope	Correlation with Y	Partial Slope
Block 1	X_1	.60	.19	.60	.19
	X_2	.60	.19	.60	.19
	X_3	.60	.19	.60	.19
Block 2	X_4	.60	.27	.60	.38
	X_5	.60	.27	.55	.13

In Set A all total correlations with Y are exactly .60. The standardized partial slopes for these data are given in the second column. Notice

that the partial slopes (.27) for the two variables in Block 2 are almost half again as large as the slopes (.19) for the three variables in Block 1. This is a reflection of the fact that there are more variables in Block 1 than in Block 2. In other words, the more variables that are used to represent any given block of highly intercorrelated variables, the less the effect attributed to any one of them. If one were to infer from this, however, that X_4 is more "important" than X_3 he would be led astray. The fewer variables one uses to represent a given block, the more important each will appear to be. In another data set, identical to the above set except for the addition of a single variable in a third block, also correlated .60 with Y, the slopes of Block 1 variables were reduced to .16 and those for Block 2 to .23, whereas the slope for the single variable in Block 3 was .41.

Suppose, now, that we introduce (as in Set B) a rather minor measurement error into X_5, resulting in a slightly reduced correlation (.55) with Y.[6] We see that the partial slope relating Y to X_5 drops sharply to .13, whereas that between Y and the other variable in Block 2 increases dramatically to .38. Gordon refers to this phenomenon as a "tipping effect", in which one (or more) variables receive credit for the explanatory power of a variable in the same block that is only slightly less strongly correlated with the dependent variable.

We can see from this example how slight attenuations produced by random measurement errors can make it extremely difficult to sort out the component effects of highly intercorrelated independent variables. Similarly, slight distortions produced by nonrandum measurement errors or sampling fluctuations could produce equally dramatic results. If one wishes to sort out these component effects, he must have both large (and

properly selected) samples and negligible measurement errors.

Choosing Among Alternative Causal Models.[7] Let us next consider the somewhat related but more general problem of choosing among causal models by a process of comparing the actual data with the predictions implied by various alternative models. Although it is conceivable that we will some day produce theories that imply specific nonzero values for certain parameters, by far the most common type of situation involves causal models that imply zero values for specific partial correlations (or slopes) owing to the assumption that there is no direct causal link between certain pairs of variables. For example, in Figure 3 there are no arrows drawn between X_1 and X_3, between X_3 and X_4, and between X_2 and X_5. This model therefore implies that the partial correlations $r_{13.2}$, $r_{34.12}$, and $r_{25.134}$ should all be approximately zero, except for sampling error. No specific predictions can be made about the remaining correlations or partial slopes, though the direct and indirect effects of each variable may be estimated from the data.

Whenever there are random or nonrandom measurement errors in some of the variables, these predictions will no longer hold, and it will be difficult to distinguish the model of Figure 3 from numerous alternatives. We can again illustrate with a very simple numerical example. Suppose we are dealing with the classical case of a spurious relationship between X and Y that is produced by a single variable W, as indicated in Figure 4. If the correlations between W and X and between W and Y are each .80, then the correlation between X and Y should be $(.80)^2$ or .64, if we assume that there are only random disturbances operating on X and Y. However, suppose all three variables are measured with random error. The

measured values of W, X, and Y are given by W', X', and Y' respectively, and let us assume that the correlation of each indicator with the true value is .80 and that the errors are purely random. (We shall later represent nonrandom errors by drawing in additional arrows linking the indicators.)

It might be thought that if all variables are measured with the same relative accuracy then our inferences would not be affected. But in multivariate models this is not the case, as can be seen if we were to attempt to partial out the effects of W by using W' instead of W. If $r_{XY} = .64$, then the partial $r_{XY \cdot W}$ (using the true values) reduces to zero since

$$r_{XY \cdot W} = \frac{r_{XY} - r_{XW}r_{YW}}{\sqrt{1 - r^2_{XW}}\sqrt{1 - r^2_{YW}}} = \frac{.64 - (.80)(.80)}{\sqrt{1 - .64}\sqrt{1 - .64}} = 0$$

However, if we use the measured values we can expect to obtain the following results (except for sampling error):

$$r_{X'Y'} = r_{X'X}r_{XW}r_{WY}r_{YY'} = (.80)^4 = .41$$

$$r_{X'W'} = r_{X'X}r_{XW}r_{WW'} = (.80)^3 = .51$$

$$r_{Y'W'} = r_{Y'Y}r_{YW}r_{WW'} = (.80)^3 = .51$$

and $$r_{X'Y' \cdot W'} = \frac{.41 - (.51)(.51)}{\sqrt{1 - .26}\sqrt{1 - .26}} = .20$$

Thus, although the true correlation between X and Y will be reduced from .64 to 0 with a control for W, the measured correlation will only be reduced from .41 to .20. One might be led to the conclusion that there is also a direct link between X and Y, or that at least there are other variables in the system that need to be controlled in order to take out the relationship between X and Y. In this very simple example, assuming that all measurement errors are purely random, it is more

important to remove the errors from the control variable W than from either X or Y. It can easily be verified that had we retained the measurement errors in X and Y, but used W instead of W' as the control, the partial correlation $r_{X'Y'.W}$ would be expected to reduce to zero, as predicted by the model.

In many practical research situations an investigator wishes to test against the possibility of a spurious relationship in order to convince a skeptic that he has, in fact, located an important cause of a given dependent variable. However, his measurement of control variables may have been much more crude than that of the variables of primary interest. Sometimes this results from the necessity of economizing at the data collection stage. In other instances the control variable may be only a very crude indicator of some other variable of greater theoretical interest. For example, one may use so-called "background variables" such as sex, age, race, father's occupation, or parental religion as indicators of socialization variables. Or it may be thought necessary to dichotomize the control variables so as to preserve enough cases in each cell. In all of these instances it may not be recognized that one is not really making a fair test of the alternative hypothesis that the relationship between X and Y is entirely spurious and due to the control variable(s) under consideration.

Nonlinearity and Nonadditivity. Simple linear additive models are of course in many instances reasonable approximations to reality, but as a science matures and improves its measuring instruments, more complex kinds of equations will be found appropriate. Furthermore, in attempting to choose among rival hypotheses, or in modifying one's

explanations so as to take alternative theories into account, it is often advisable to utilize more complex models than simple linear additive ones. Obviously, if there are extensive random or nonrandom measurement errors that produce "noise" in the data, it becomes difficult to introduce these needed refinements.

Perhaps a few examples will be sufficient to illustrate this rather obvious point. Many of our verbal propositions in sociology take the qualitative form "the greater the X, the greater the Y." Aside from certain ambiguities as to causal symmetry or asymmetry involved in such "greater-greater" statements, there is usually nothing said about the exact form of the relationship implied.[8] Perhaps linearity is intended, but often there may be a kind of "saturation effect" or "diminishing returns" argument implicit in the discussion of the theory. In other words, it may not be expected that equal increments of X will produce the same changes in Y as the level of X increases. In such an instance, it would be preferable to state the proposition as follows: "Y is a monotonic increasing function of X that has a decreasing slope." This would imply a nonlinear relationship of the form given in Figure 5, which might be represented by an equation expressing Y as a positive function of logX.

If we wished to test such a proposition, we would need to consider various kinds of artifacts produced by measurement error that might also generate the same kind of curve. One obvious possibility is that the instrument used to measure Y is relatively insensitive to differences among high values, so that individuals with high Y scores are all bunched together near the upper extreme of the measured Y scores, even where they differ considerably on their true Y scores. For example, a "political conservatism" scale may lump together a wide range of individuals simply

because many persons of varying persuasions have endorsed nearly all of the items. Conceivably there may be relatively more random measurement error at one end of the Y continuum than the other, perhaps because of a greater illiteracy at that extreme. Or rapport and understanding of the research objectives may be poorer at one extreme than the other, producing a regression toward the mean that is more extensive in one direction than the other.

Similarly, measurement errors can produce artifacts leading to statistical interaction or nonadditivity in the data.[9] It has already been noted that purely random measurement errors will attenuate slope estimates in proportion to the ratio $\sigma^2_X/\sigma^2_{X'}$. Recalling that in the case of random errors $\sigma^2_{X'} = \sigma^2_X + \sigma^2_e$, we see that this implies that even where σ^2_e remains constant, we may have differential attenuation according to the dispersion in the true X scores. Likewise, if there happens to be a greater measurement error component σ^2_e for one subpopulation than another, the measured slopes may differ even though the true slopes are equal. This will produce statistical interaction, as tested by analysis covariance. A similar phenomenon will occur in the case of categorized data. If contingency tables are used, relationships will appear weaker in those tables representing subpopulations with the greatest relative random measurement errors.

There may also be nonrandom measurement errors produced by crude classifications. Suppose, for example, that workers have been dichotomized into white- and blue-collar occupations and as having "high" and "low" educations. This will produce four cells, but it is erroneous to assume that either occupation or education is being "held constant" in any column or row. Persons who are "high" on education are likely to have

very different white-collar occupations than are white-collar workers with "low" educations. If the joint effects of these two variables on some dependent variable are being studied, one may infer nonadditive effects when an additive model would be more realistic.

For example, a status-inconsistency theory might predict that persons who are either low-high or high-low will be more liberal than those who are high-high or low-low. This prediction would show up as a difference of sums, as in the following table:

		Education	
		High	Low
Occupation	White-collar	30	60
	Blue-collar	60	70

where high scores in the body of the table represent high liberalism scores.[10] Thus the sum of the HH and LL cells is 100, whereas that for the inconsistent cells, LH and HL, is 120. But an alternative explanation might be that high occupation, alone, produces conservatism. The very low liberalism scores in the top left cell might merely reflect the possibility that the white-collar persons in this cell are primarily professionals and managers, whereas those in the bottom left cell are in the clerical and sales occupations. Similarly, persons classed as "blue-collar high education" and those as "white-collar low education" might have occupations of very similar status.

It is hoped that the several kinds of illustrations of distortions that can be produced by errors of measurement are sufficient to convince the reader that, if we are to improve our testing and estimating procedures, we must simultaneously pay more attention to problems of measurement. As long as exploratory studies involving large numbers of

variables are in vogue, measurement errors may not prove insurmountable barriers. But as soon as we begin to take seriously the tasks of refining our theories and of testing them against the numerous possible competing alternatives, we will very quickly reach an impasse without improved measurement.

In the next section we shall consider various statistical approaches to estimating the extent of measurement error and correcting for its presence. We shall see that these procedures require relatively strong a priori theoretical assumptions unless our assumptions about measurement errors can be kept relatively simple. Of course it is preferable to purify measures at the data collection stage, rather than attempting to remove errors at the analysis stage. Nevertheless, a careful study of just what can and cannot be accomplished after the data have been collected should provide guidelines and cautions with important implications.

THE STATISTICAL ESTIMATION OF RANDOM AND NONRANDOM MEASUREMENT ERRORS

Random Errors. In the case of purely random measurement errors, it turns out that random errors in a dependent variable Y will not systematically bias a slope estimate, though they will attenuate correlations with Y. But as we have seen, random errors in an independent variable X will attenuate correlations and slopes, the latter according to the approximate formula

$$E(b_{YX'}) = \frac{\beta_{YX}}{1 + \sigma^2_e/\sigma^2_X}$$

One obvious way to correct for this bias is to insert estimates of σ^2_e and σ^2_X, or their ratio. But how can these be obtained? Not from the data themselves. In sociology, at least at the present time, it is

difficult to imagine how we would obtain the necessary information without a good deal more standardization and careful assessment of data collection techniques.

A second approach that would be extremely handy if it actually worked is to utilize certain grouping procedures suggested by Wald and Bartlett, among others.[11] These essentially involve ranking individuals according to their X' scores, placing them into two groups according to these X' scores (say, group 1 = top third and group 2 = bottom third), taking mean X' and Y scores for each group $(\overline{X}'_1, \overline{Y}_1)$ and $(\overline{X}'_2, \overline{Y}_2)$ and then forming the simple slope estimate

$$\tilde{b}_{YX} = \frac{\overline{Y}_2 - \overline{Y}_1}{\overline{X}'_2 - \overline{X}'_1}$$

Wald has shown that, if the method of grouping is independent of the measurement error in X, then the proposed estimator will have negligible biases. But in practical applications it appears to be impossible to meet this critical assumption, since grouping must be done by the X' scores, rather than the true X values. Several sets of computer-generated data have led to the conclusion that the estimated biases using the Wald-Bartlett procedure are practically identical to those obtained with ordinary least squares.[12] It therefore appears as though this very simple attack on the problem will not prove fruitful, although additional studies will be needed to demonstrate this convincingly.

A third approach, which holds considerably more promise provided we can gain greater confidence in our theories, involves the use of what have been termed "instrumental variables."[13] In brief, this approach requires us to locate variables that are direct or indirect causes

of the independent variable X but that do not appear in the equation for the dependent variable Y. For example, in the case of the simple model $W \rightarrow X \rightarrow Y$ the instrumental-variable estimator of β_{YX}, denoted by $b^*_{YX'}$, involves the ratio of the (sample) covariances of Y and X' with the instrumental variable W. Thus

$$b^*_{YX'} = \frac{\Sigma wy}{\Sigma wx'}$$

where the small-case letters refer to deviations around the respective means, and where we again assume that $X' = X + e$, with e being a random variable.

It can be shown that $b^*_{YX'}$ is a consistent estimator of β_{YX}, which means that its large-sample biases will be negligible even with rather large random measurement errors in X. We have shown, using computer-generated data, that if the assumptions of the model are met, the instrumental-variables estimator works much better than ordinary least squares in the presence of varying amounts of random measurement errors.[14]

However, if there is "specification error," or error in the model itself, the instrumental-variables approach is likely to produce greater biases (and also greater standard errors) than ordinary least squares. In the important case where W appears in the equation for Y, contrary to assumption, it can be shown that <u>both</u> ordinary least squares and the instrumental-variables estimators will involve biases, apart from the question of measurement error in X. To the degree that W and X are highly correlated, these biases will be approximately equal, but the weaker the relationship between W and X, the greater the relative bias of the instrumental-variables estimator. If W does appear in the equation for Y, this means that W causes Y through one or more paths in addition to

the path through X. Of course if these alternative paths can be accounted for, say by introducing additional intervening links explicitly into the equation for Y, then this particular problem can be handled. We see, however, that this approach requires that one make rather strong theoretical assumptions (about the operation of a third variable) in order to assess measurement error.

The final two approaches that will be summarized in this section both require multiple measures of X. In the first of these, there must ordinarily be two or more measures of each variable that has been imperfectly measured. Although this general approach is also utilized in factor analysis and in corrections for attenuation commonly made in psychological testing, we shall illustrate an explicitly causal version due to Costner.[15] A simple model exemplifying this approach is given in Figure 6. The two indicators of X are designated as X_1 and X_2, and similarly for Y. The absence of additional arrows connecting the indicators implies that the measurement errors in all cases are assumed to be completely random. The letters a, b, c, d, and e represent path coefficients, which in this particular model are also (unmeasured) correlation coefficients. We allow for the possibility that the measures of both X and Y are not equally good by not imposing the restriction that a = b or d = e, as will later be done in the case of overtime data.

Since X and Y are themselves unmeasured, none of the path coefficients can be obtained in a direct fashion. However, there will be six intercorrelations among the four indicators X_1, X_2, Y_1, and Y_2, and we may write an equation for each of these correlations in terms of the five unknown path coefficients. These are as follows:

$$r_{X_1X_2} = ab \qquad r_{X_1Y_2} = ace$$
$$r_{Y_1Y_2} = de \qquad r_{X_2Y_1} = bcd$$
$$r_{X_1Y_1} = acd \qquad r_{X_2Y_2} = bce$$

It can be seen that we can then solve for each of the path coefficients. For example

$$a^2 = \frac{a^2bcd}{bcd} = \frac{r_{X_1X_2}r_{X_1Y_1}}{r_{X_2Y_1}} \text{ and } c^2 = \frac{abc^2de}{abde} = \frac{r_{X_1Y_1}r_{X_2Y_2}}{r_{X_1X_2}r_{Y_1Y_2}}$$

Also, since there is one more equation than unknown, we obtain a redundant equation that can be used to test the adequacy of the model, since it will not automatically be satisfied by all data. One way of expressing this equation is as follows:

$$r_{X_1Y_1}r_{X_2Y_2} = r_{X_1Y_2}r_{X_2Y_1} = abc^2de$$

This can be thought of as a generalization of the procedure used in correcting for attenuation, where it is additionally assumed that so-called equivalent forms have been used, so that we may set a = b and d = e. It should be cautioned, however, that certain specific models involving nonrandom measurement error can also satisfy the above condition. One such possibility will be discussed below; others are considered by Costner. Therefore this condition can be considered as necessary but not sufficient. Of course in real instances there will be sampling error superimposed on measurement error, so that the condition will never be exactly satisfied, even where measurement errors are completely random. The point is that if the assumptions of the model seem reasonable, we may use the correlations among indicators to infer all of the true correlations or path coefficients.

Heise discusses a very similar approach that can be used whenever

one has single indicators of variables at three or more points in time.[16] In the model of Figure 7 there are two indicators X_{11} and X_{i2} at each of three times (i = 1, 2, 3), but we shall for the time being consider only the first indicator at each time period. The model assumes that there is the same relative measurement error σ^2_e/σ^2_X at each time period. This assumption permits us to use the same path coefficient a representing the link between X and its indicator, thereby reducing the number of unknowns. If one uses the same measuring instrument at each time it might be plausible to assume the constancy of σ^2_e, though of course the variance in X will depend on the effects of extraneous factors that may not continue to operate so as to justify this important simplifying assumption.

If we first assume that data involving the single indicator were available at only times 1 and 2, we would have only a single correlation with which to estimate the two coefficients a and c. Thus

$$r_{X_{11}X_{21}} = a^2c$$

If we were willing to assume that X remains constant over the interval, this would imply that c = 1, and we could therefore obtain an estimate of a^2 as an indication of measurement reliability. But if there is true change in X, plus random measurement error, we cannot estimate either coefficient unless we have a third observation and unless we make additional assumptions, such as those implied in the model of Figure 7. With three points in time we have three equations and three unknowns as follows:

$$r_{X_{11}X_{21}} = a^2c \qquad \text{and} \qquad r_{X_{21}X_{31}} = a^2d$$

$$r_{X_{11}X_{31}} = a^2cd$$

and we see that the path coefficients can all be estimated. Thus

$$a^2 = \frac{r_{X_{11}X_{21}} r_{X_{21}X_{31}}}{r_{X_{11}X_{31}}}, \quad c = \frac{r_{X_{11}X_{31}}}{r_{X_{21}X_{31}}}, \text{ and } d = \frac{r_{X_{11}X_{31}}}{r_{X_{11}X_{21}}}$$

These procedures suggested by Costner and Heise can be combined rather simply and, as we shall see, such a combination can be used effectively to infer certain kinds of nonrandom errors.[17] In the case of purely random errors, we can see that we can use two measures of X at only two points in time to infer the coefficients. Thus in Figure 7 if we utilize the measures X_{11}, X_{12}, X_{21}, and X_{22} we have a situation that is identical to the model of Figure 6, with the added simplifications made possible by the fact that Y has been replaced by X at time 2. These simplifications reduce the number of unknowns by two and provide two additional equations that must be satisfied by the data if the model is to be retained. If we use more than two indicators of X, or more than the two time periods, we obtain further redundant equations for testing purposes.

Nonrandom Errors. When we introduce the realistic possibility that measurement errors may be nonrandom, we open Pandora's box. On the conceptual level, we must face up to the problem of developing auxiliary theories specifying our assumptions as to the linkages between measured and unmeasured variables. Where a single indicator is linked to several unmeasured variables, the model is likely to become too complex to handle without additional simplifications. In general, the higher the percentage of unmeasured variables in the system, the simpler the model must be in other respects. This point should become more apparent as we proceed.

Perhaps the simplest kind of nonrandom measurement error is a

constant error that affects only intercepts in a regression equation. Somewhat less simple are sources of measurement error in X that can be taken as functions of the level of X itself. In these instances we may again write $X' = X + e$, where e is some function of X. Hopefully, it may be reasonable to take the error component as a linear function of X, plus a completely random component u, so that we may write an equation for X' in the form $X' = a + bX + u$, where u is unrelated to all variables in the system. It can be shown in this instance that Costner's procedure for estimating path coefficients (or correlations) can be applied exactly as before, although the true slope connecting Y to X cannot be estimated because of the error in scale appearing in X'.[18]

Errors proportional to X can arise as a result of several kinds of distortions. For example, there may be a regression toward the mean of X in an attitudinal questionnaire produced by a norm favoring moderate or intermediate responses. That is, extreme liberals may tend to answer as moderately liberal, and extreme conservatives as moderately conservative. There may also be errors due to "ceiling" or "floor" effects that are primarily in the direction of less extreme scores.[19] Whenever a small number of ordered categories are used, extreme true scores are likely to be bunched together, with the seriousness of the distortions being a function of the number of categories and the relative numbers of individuals appearing in the extreme categories. Although these kinds of measurement errors can best be described in terms of nonlinear functions of X, a linear function with a negative slope may be taken as a reasonable approximation.

Other kinds of nonrandom errors are not so easy to handle. In

many instances we expect nonrandom errors to be produced by extraneous factors that are difficult to identify and measure. For example, governments may distort their official records in systematic ways, so as to improve their outside images. Individuals may similarly tend to give conventional answers on many different kinds of questions. Interviewer biases may persist across many different sets of items, and so forth. In all of these instances, our first task (ideally) is to construct a model containing the presumed sources of nonrandom error linked to whichever system variables seems appropriate. If these disturbances can be measured and their effects thereby taken into consideration, there may be no special difficulties. But if they cannot, it will be only under very special circumstances that their effects can be inferred. For example, if one has more than two indicators of each variable, and if only a few indicators are subject to such nonrandom disturbances, Costner has shown that these effects can sometimes be estimated.[20]

Suppose, however, there are only two measures of X both of which are affected by Z, which may also be linked to X and the other variables in the system. One such possibility can be diagrammed as in Figure 8, where Z is taken as a cause of W, X, and Y. It can be shown that if both X and Z are taken as unmeasured, it will be impossible to estimate the path coefficients (correlations) between W and X and between X and Y unless Z is unrelated to <u>all three</u> variables W, X, and Y.[21] Furthermore, even under these ideal conditions it is necessary to assume the simple causal chain $W \rightarrow X \rightarrow Y$ in order to estimate the link between X and Y. In other words, we must find an instrumental variable W, related to the other variables in a very simple way, before we can

estimate the true correlation between X and Y in this model in which Z is assumed to affect both measures of X.

If we are fortunate enough to have two measures of each variable at three points in time, we may handle a larger variety of nonrandom measurement errors. Let us consider a single variable X and the model of Figure 9. This model allows for a simultaneous disturbance path f between the two indicators at each point in time; equal disturbances g between the first indicator between times 1 and 2, and between times 2 and 3; a different disturbance g' between times 1 and 3; and similar disturbance terms h and h' for the second indicator. Disturbances that might affect relationships between different indicators at different times, however, are ruled out. This model contains six measured variables, and therefore there will be fifteen equations to estimate the nine unknowns.

It is not always true that if there are excess equations a solution can be obtained, since most of the equations may be redundant. However in this instance we may solve for the unknowns. The coefficients c and d, which measure the stability of X over time, can be estimated by the simple equations

$$c = \frac{r_{x_{11}x_{32}}}{r_{x_{21}x_{32}}} = \frac{abcd}{abd} \text{ and } d = \frac{r_{x_{11}x_{32}}}{r_{x_{11}x_{22}}} = \frac{abcd}{abc}$$

Some of the remaining coefficients, however, can only be estimated reliably under special conditions. For example, the estimates of a^2 and b^2 involve the difference between c and d, or the difference in stability between times 1 and 2 and between times 2 and 3. Unless c - d is relatively large, these estimates may involve extremely large sampling errors.[22]

The practical implication suggested by these few examples -- and

there are many other possibilities that need to be studied--is that the more complex and realistic we make our assumptions about measurement errors, the more unknowns we introduce into the picture, and the greater the price we must pay in other respects. We must either include other instrumental variables in such a way that they are simply related to the basic variables of interest, or we must have multiple measures of each unmeasured variable. The picture becomes even more disturbing when we consider theories containing variables that are only very indirectly measured, and where we know that there are complex connections between underlying variables and their indicators.

For example, consider the measurement of several different kinds of minority discrimination. If discrimination is defined theoretically as the differential treatment of minority members with respect to educational opportunities, housing, occupation, income, and so forth, then we will seldom be in a position to observe such discrimination directly. Instead, our measures will consist of inequalities or degree of segregation. It would be convenient to take educational inequalities as indicators of job discrimination, and income inequalities as indicators of income discrimination, but this would indeed be too simple. Not only would this neglect differences in motivation, other forms of minority behavior, and possible innate differences, but it would ignore the fact that income inequalities depend on educational discrimination, occupational discrimination, and income discrimination. Residential segregation may also depend on all three, and so forth. Therefore we will have multiple indicators of multiple unmeasured variables, and there will be far too many unknowns for solution.

Under certain limited kinds of restrictive assumptions, multiple

factor analysis may be used to help disentangle the variables. But when some of the indicators are also linked by additional variables, the situation becomes much more difficult to handle. Suffice it to say that we have hardly begun to explore the implications of these kinds of measurement complications. All too often, our simplistic assumptions are not made explicit. For example, in the above illustration if we were to proceed by taking educational inequality as the indicator of educational discrimination, and so forth, this would amount to assuming no direct links between the remaining types of discrimination and the other indicators. It would seem far better to force such assumptions out into the open, where they can be subjected to careful scrutiny, than to hide them from view. For unless they are made fully explicit we cannot begin the difficult task of analyzing our measurement procedures.

IMPLICATIONS

Of the two major obstacles mentioned in the introduction, I am relatively optimistic that we can proceed rather systematically to build more and more complex theories that are increasingly realistic. Once we have formed the habit of stating our propositions and assumptions explicitly, and once we have learned to tolerate theoretical models that are relatively more simple than we might like, we will set in motion a cumulative process through which inadequacies in each formulation can gradually be corrected. Of course we must avoid the temptation of introducing so many modifications that the theory becomes inherently untestable because of the presence of too many unknowns in the system. But there seem to be few methodological or psychological reasons why we cannot learn to construct highly complex theories that

are, in principle, testable.

The major roadblock, as has been implied throughout this discussion, would appear to consist of a series of problems associated with the measurement process. These problems are both methodological-conceptual and practical. Data collection is both expensive and time-consuming, and unfortunately sociologists can seldom rely on outside agencies to collect data on variables of primary interest. Although we may make heavy use of census data and other kinds of official records, we usually find that we must postulate rather tenuous links between the indicators that are most readily available and the conceptual variables appearing in our theories. As we become increasingly interested in cross-national studies, these practical problems will be all the more serious. It seems safe to assume, therefore, that data collection will lag considerably behind both our methodological sophistication and our ability to formulate theoretical models.

I believe this will have serious implications for the profession and the way we are organized to conduct our research. At present, we are poorly coordinated. Individual or small team research is the prevailing pattern, and few attempts have been made to consolidate and coordinate studies, to standardize variables and their indicators, or to replicate research in systematic ways. Greater prestige is awarded to the person who constructs a new scale than to the investigator who patiently replicates and revises an older one. We appear to cling to the belief that superior measuring instruments and concepts will eventually win out in the competition with the others, but we have hardly begun even to define the rules under which the footrace should be run. As a result, the number of concepts and indicators seems to be

proliferating at an alarming rate.

In spite of the obvious disadvantages it would entail, I believe it is time to begin the effort to coordinate our activities, to standardize our measuring instruments, and to expand the scale of our research operations. This cannot be done in such a way as to inhibit flexibility or to discourage exploratory research by single investigators. Obviously a division of labor is necessary, particularly in connection with data collection procedures. Perhaps it will be possible to select one or two subfields on an experimental basis, to form data-collection institutes in these substantive areas, and to formulate rather ambitious projects that make it possible to collect longitudinal as well as cross-national data.

The dangers of political control and of elitism are rather obvious in such operations, and for this reason it may be necesaary to assure that boards of social scientists control the basic policies of such institutes and that all data be made available to individual scholars at a reasonable cost. At the same time, careful quality control must be instituted. An advantage of the large-scale institute is that it can afford to hire measurement specialists who may conduct methodological studies and attempt to assess the extent and nature of measurement error in each variable. Hopefully, if a number of individual analysts all have access to the same data sets, and if they can agree on standard terminology, we may begin the long road toward reducing the number of slightly different vocabulary systems that presently exist in the field.

The reward structure of the profession must also be changed to a major extent. At present we tend to reward either the quick study

that can be immediately worked into a journal article or the book-length monograph that contains table after table, analyzed in such a way that the intelligent layman can follow the argument without difficulty. Such a system does not motivate one to conduct longitudinal studies of more than a year or two in duration or to utilize more sophisticated methods of data analysis. Since data collection by individuals is very expensive--unless it is done on the college campus--replication studies have low payoff. Furthermore, they are harder to get published in the major journals. Hopefully, the presence of a number of data-collection institutes with reasonably long life span could help correct this deficiency, particularly if secondary data were readily made available to graduate students for term papers and M.A. theses.

Although it is difficult to know how to correct for the problem, it seems obvious that, from the standpoint of the necessity of improving our measurement procedures, sociologists are spreading themselves too thin. We cannot possibly study <u>intensively</u> every interesting social phenomenon that might conceivably be "relevant" to sociology, and yet it would be unwise to attempt to eliminate substantive areas by fiat, if this were even possible. Again, given the shortage of manpower and the fact that sociologists are heavily engaged in the teaching function as well as in research, it would seem wise to attempt to select a few substantive areas for intensive research. Presumably, the methodological problems encountered in these areas will be sufficiently similar to those areas less intensively studied that they can be applied much more generally.

It is not easy to suggest a definite list of substantive areas with

high potential, or to formulate a practical mechanism by which selective concentration can be brought about. Clearly, some social problem areas are capable of attracting more research funds than others, but it does not follow that forced feeding will produce important results. For example, the field of "medical sociology" has enjoyed large-scale federal financing in the U.S., but the field appears to be too diffuse and too low prestige (at least in America) to have had the desired theoretical and methodological payoff. Fields that are closer to the "core" of sociology, such as stratification, large-scale organizations, social psychology, and human ecology, do not seem to have attracted major funds, whereas certain others, such as race relations and population, have been subjected to "on-and-off" policy decisions that have made long-range planning difficult.

Partly because of the fact that specialists with common interests are scattered in different educational institutions in order to attain balance with respect to the teaching function, sociologists in a single substantive field are seldom able to communicate extensively except at annual meetings or brief conferences. It therefore seems advisable for the profession to obtain funds to bring such specialists together for prolonged periods for the purpose of cross-fertilization and standardization.

Finally, we very much need to correct various imbalances in the degree to which different kinds of methodological problems are taken into consideration in our research decisions. Of course this is partly a function of differential ignorance of the consequences of errors of various kinds. Thus we seem to be much more aware of the properties of different levels of measurement (e.g., ordinal versus

interval scales) than of different kinds of nonrandom measurement errors. Similarly, we are much better acquainted with scaling procedures and ways of inferring multidimensionality than with problems of aggregation and autocorrelation that have been studied by econometricians. To the degree that one is a methodological "purist" who strives to avoid making unreasonable assumptions, this kind of differential knowledge of consequences of unmet assumptions may have a considerable impact on analysis strategies. For example, I am convinced that one of the reasons why survey data are often "under-analyzed" is that many investigators, wishing to avoid dubious assumptions to the effect that an interval-scale level of measurement has been attained, resort to simple cross-tabulations involving at most three or four variables. Not only does such a practice amplify measurement errors but, in effect, it forces one to assume that a very small subset of variables can be analyzed apart from the rest.

In short, much remains to be accomplished before we can begin to "test" our theories in a satisfactory way. It would be unwise, however, to become so overly purist that we are immobilized because of well-recognized but unmet assumptions. It is best to plunge ahead, making our assumptions explicit and examining their implications one by one. But it is also time to include our assumptions about measurement errors and to begin the very important task of reorganizing our research operations so as to close the gap between theory and research.

FOOTNOTES

1. For a discussion of the role of such auxiliary theories see H. M. Blalock, "The Measurement Problem: A Gap Between the Languages of Theory and Research," in H. M. Blalock and Ann B. Blalock (eds.), Methodology in Social Research (New York: McGraw-Hill, 1968), Chap. 1.

2. For a technical discussion of the properties of block-recursive systems see Franklin M. Fisher, The Identification Problem in Econometrics (New York: McGraw-Hill, 1966), Chap. 4. For a much less technical discussion see H. M. Blalock, Theory Construction (Englewood Cliffs, N.J.: Prentice-Hall, 1969), Chap. 4.

3. Fisher, op. cit., p. 101.

4. See J. Johnston, Econometric Methods (New York: McGraw-Hill, 1963), pp. 149-150.

5. Robert A. Gordon, "Issues in Multiple Regression," American Journal of Sociology, 73 (March 1968), pp. 592-616.

6. Such measurement error would of course also reduce correlations with the other independent variables, so that minor changes in Gordon's figures would be required in a more realistic example.

7. For a more complete discussion of these and other kinds of distortions see H. M. Blalock, "Some Implications of Random Measurement Error for Causal Inferences," American Journal of Sociology, 71 (July 1965), pp. 37-47.

8. For discussions of the causal asymmetry problem in this connection see Herbert L. Costner and Robert K. Leik, "Deductions from 'Axiomatic Theory'," American Sociological Review, 29 (December 1964), pp. 819-835; and Blalock, Theory Construction, op. cit., Chap. 2.

9. Additional examples of this type are discussed in H. M. Blalock, "Tests of Status Inconsistency Theory: A Note of Caution," Pacific Sociological Review, 10 (Fall 1967), pp. 69-74.

10. See Gerhard E. Lenski, "Comment," Public Opinion Quarterly, 28 (Summer 1964), pp. 326-330.

11. See Abraham Wald, "The Fitting of Straight Lines if Both Variables are Subject to Error," Annals of Mathematical Statistics, 2 (1940), pp. 284-300; M. S. Bartlett, "Fitting a Straight Line when Both Variables are Subject to Error," Biometrics, 5 (June 1949), pp. 207-212; and Albert Madansky, "The Fitting of Straight Lines when Both Variables are Subject to Error," Journal of the American Statistical Association, 54 (March 1959), pp. 173-205.

12. H. M. Blalock, Caryll S. Wells, and Lewis F. Carter, "The Statistical Estimation of Random Measurement Error" (unpublished manuscript).

13. For discussions of instrumental variables see Johnston, op. cit., pp. 165-166; Carl Christ, Econometric Models and Methods (New York; John Wiley, 1966), pp. 404-410; and Blalock, Wells, and Carter, op. cit.

14. Ibid.

15. See Herbert L. Costner, "Theory, Deduction and Rules of Correspondence," American Journal of Sociology, 75 (September 1969, pp.

16. See David R. Heise, "Separating Reliability and Stability in Test-Retest Correlation," American Sociological Review, 34 (February 1969), pp. 93-101.

17. See H. M. Blalock, "Estimating Measurement Error using Multiple Indicators and Several Points in Time" (unpublished manuscript).

18. See H. M. Blalock, "A Causal Approach to Nonrandom Measurement

Errors" (unpublished manuscript).

19. For further discussion of ceiling effects see Paul M. Siegel and Robert W. Hodge, "A Causal Approach to the Study of Measurement Error," in Blalock and Blalock, op. cit., Chap. 2.

20. See Costner, op. cit.

21. Blalock, "A Causal Approach to Nonrandom Measurement Errors," op. cit.

22. Blalock, "Estimating Measurement Error," op. cit.

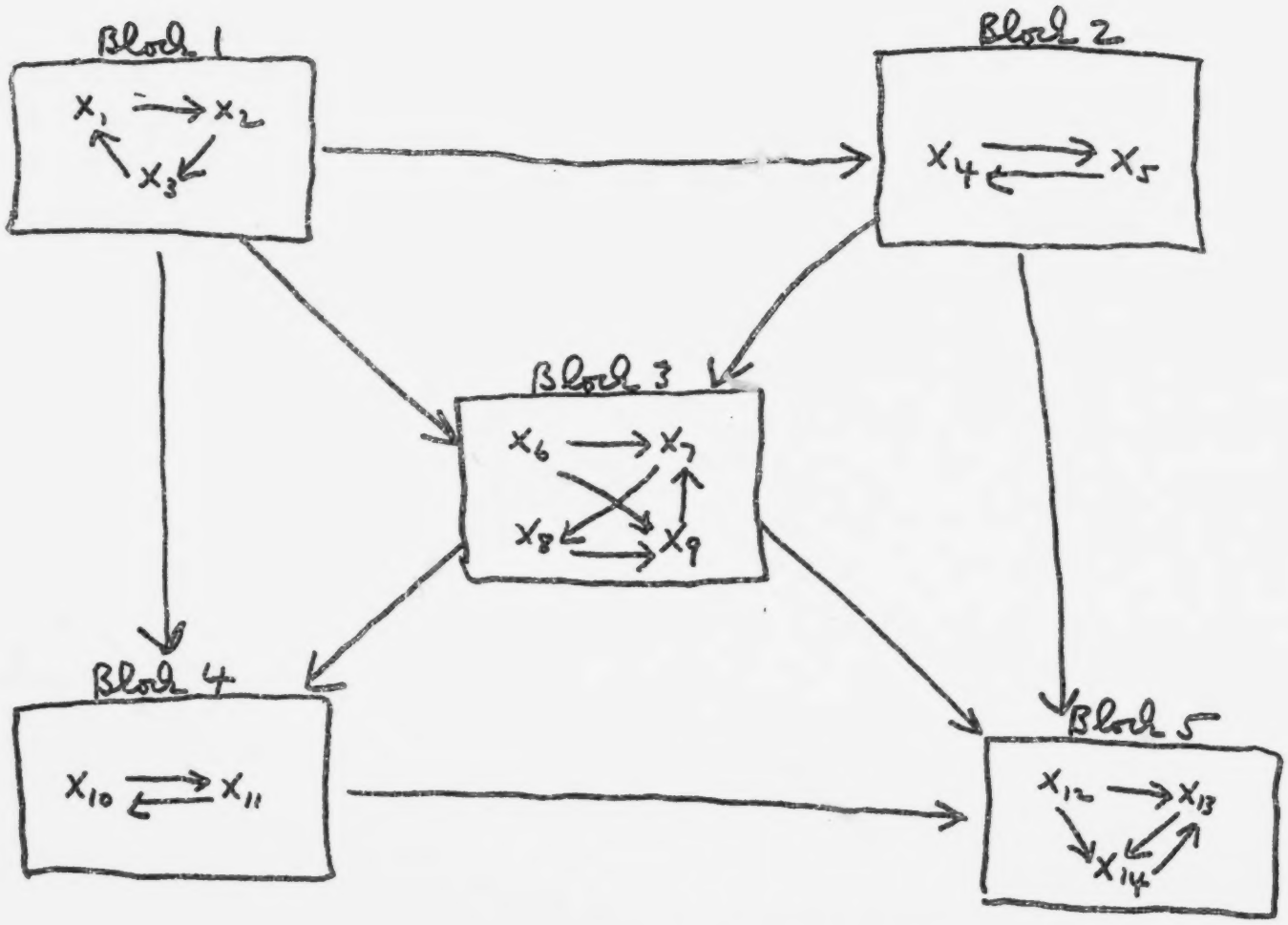


Figure 1.

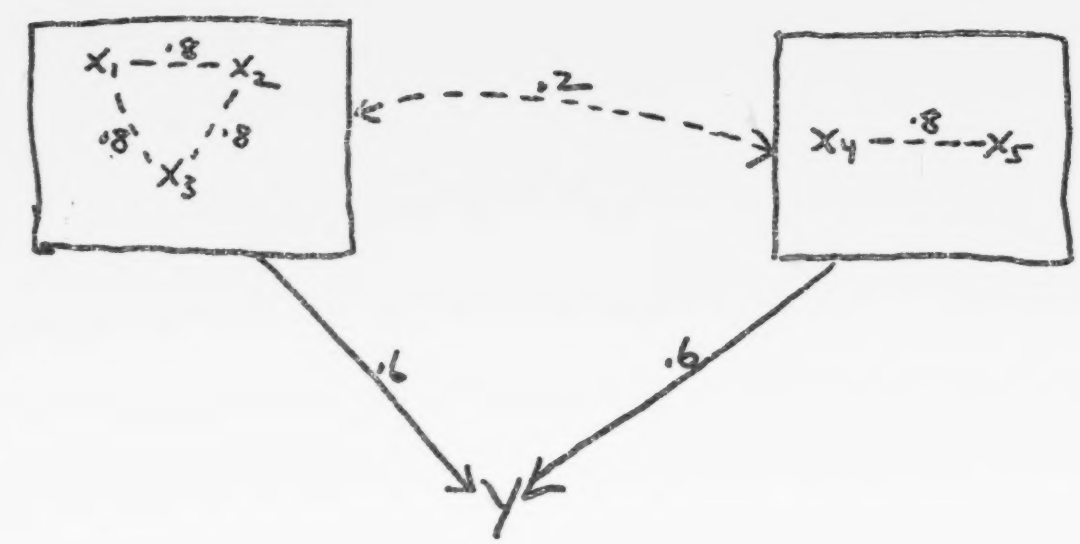
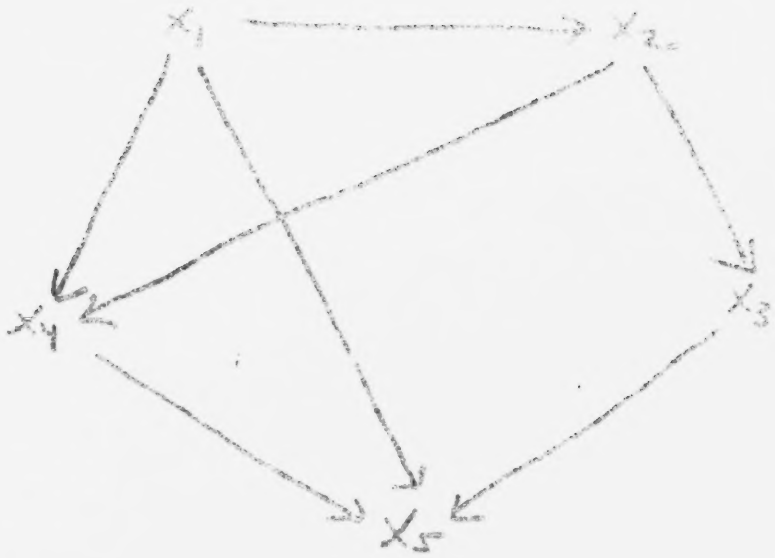


Figure 2.



$\rho_{13.2} = 0$

$\rho_{34.12} = 0$

$\rho_{25.134} = 0$

Figure 3.

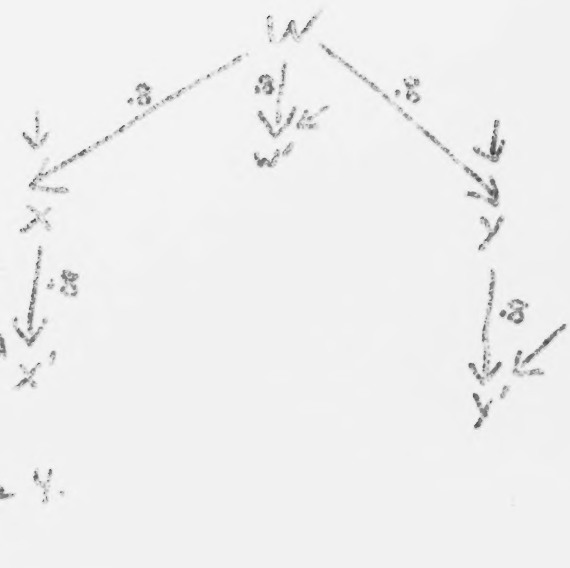


Figure 4.

Figure 5.

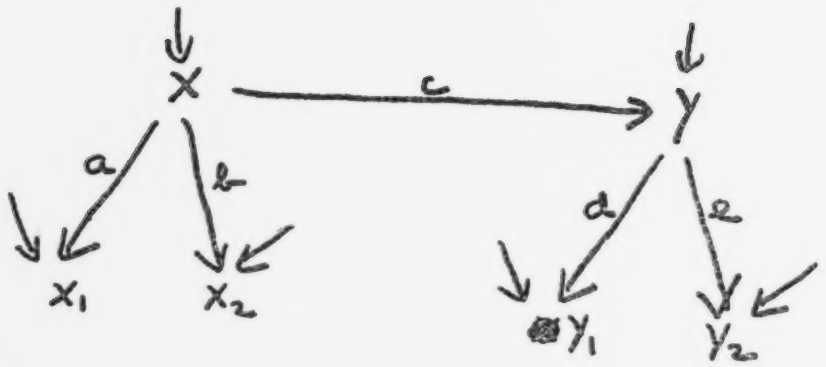


Figure 6.

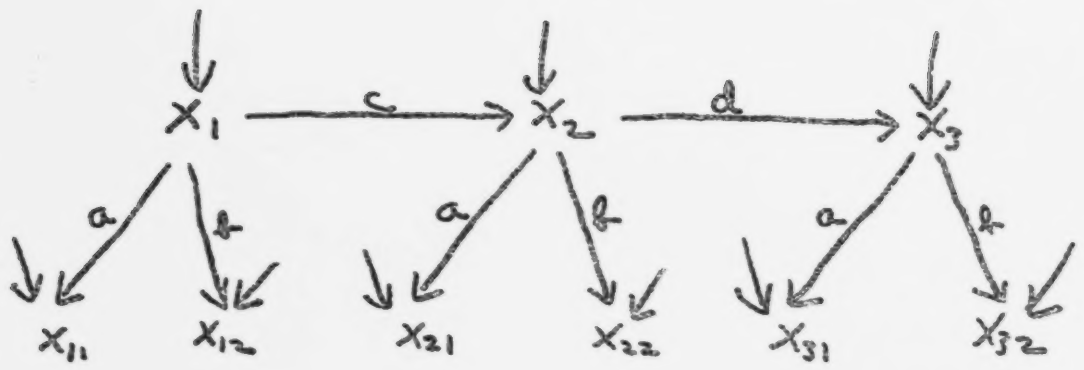


Figure 7.

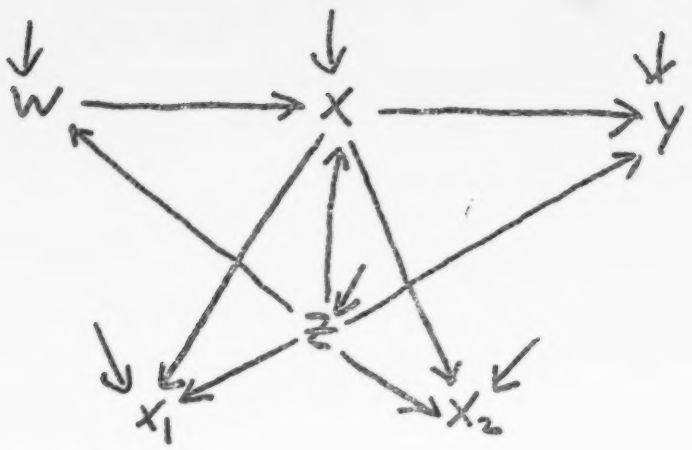


Figure 8.

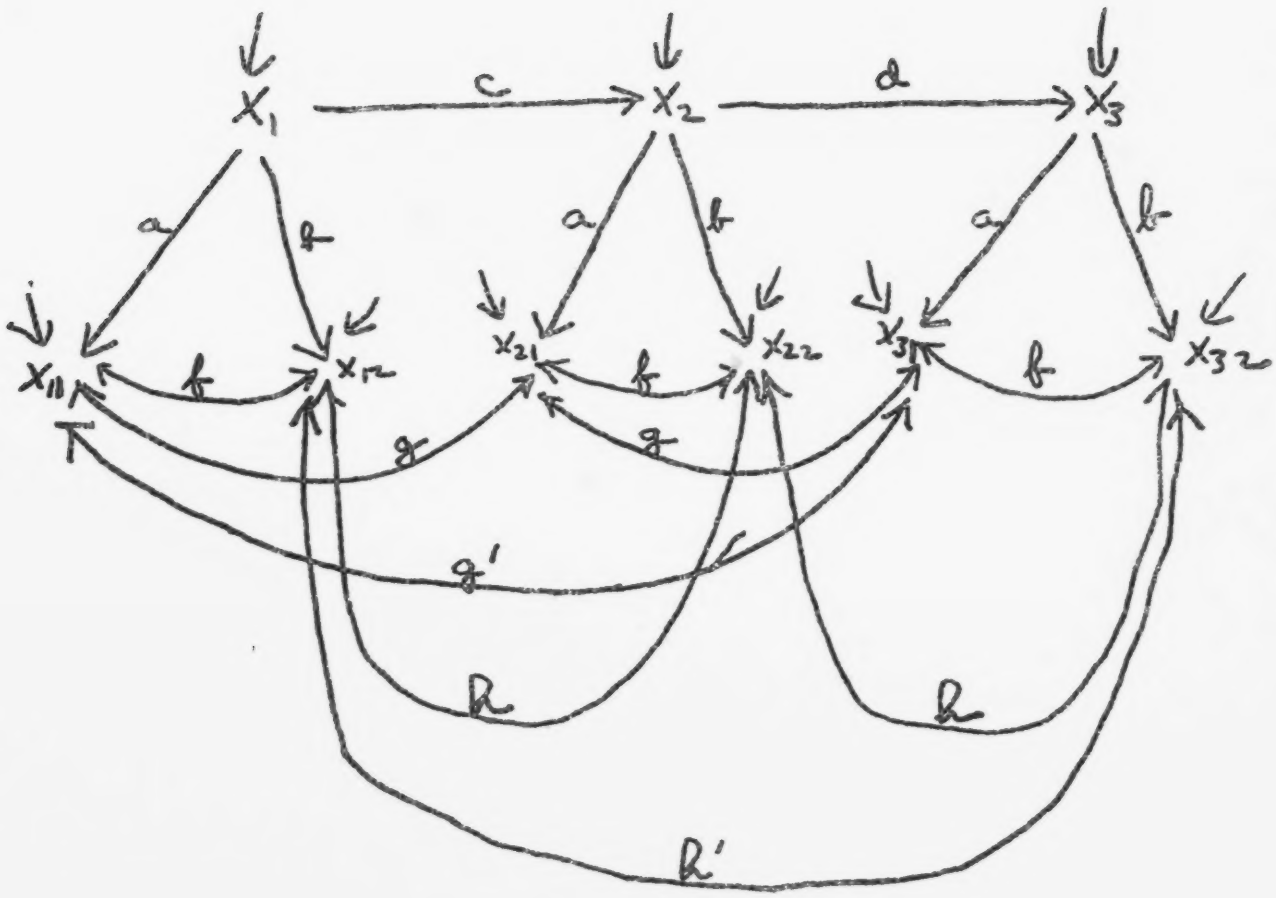


Figure 9.